# Multi-parametric statistical analysis of economic data for continuous pharmaceutical manufacturing

Zdravko Kravanja (Editor), Proceedings of the 26th European Symposium on Computer-Aided Process Engineering – ESCAPE 26
June 12th -15th, 2016, Portorož, Slovenia 

# Multi-parametric Statistical Analysis of Economic Data for Continuous Pharmaceutical Manufacturing

Brigitta Nagy,[a] György Marosi,[a] Dimitrios I. Gerogiorgis[b]

[a] *Department of Organic Chemistry & Technology, Budapest University of Technology & Economics, Megyetem rkp. 3, Budapest H-1111, Hungary*
[b] *School of Engineering (IMP), University of Edinburgh, Edinburgh EH9 3FB, UK*

## Abstract

The global pharmaceutical industry faces high R&D, regulatory and cost pressure which can be alleviated by the advent of Continuous Pharmaceutical Manufacturing (CPM). Embarking upon demonstrating and commissioning continuous processes is not trivial: judicious product selection and process design is quintessential for viable investments. Technological as well as economic considerations must be inseparably combined, but a quantitative method for elucidating only the most viable candidates has yet to emerge. This study illustrates how systematic statistical analysis can support business decisions and process R&D for the synthesis and design of continuous pharmaceutical processes. A systematic statistical evaluation of UK economic data has been performed to identify viable drug substances (DS) and drug products (DP) for continuous manufacturing. Product classification and ranking is employed to select those with the highest demand, and statistical hypothesis testing explores causality and correlations of key parameters. Molecular weight and complexity have been correlated with trade and value statistics, indicating that amides, lactones, antibiotics and hormones have high CPM potential.

**Keywords**: Continuous Pharmaceutical Manufacturing (CPM), economics, statistics.

## 1. Introduction

The pharmaceutical industry faces many pressing R&D, quality, supply chain and cost challenges due to the steadily decreasing drug discoveries and the increasing market share of generics manufacturers which diminishes profitability margins (Plumb, 2005). Continuous Pharmaceutical Manufacturing (CPM) emerges as a viable opportunity to simultaneously streamline product development and improve process economics: it promises a potential competitive edge to pharmaceutical corporations by capital and operational expenditure savings due to smaller equipment units, reduced footprint and enhanced process understanding and efficiency (Schaber et al., 2011; Anderson, 2012). Successful CPM implementation critically depends on technical advances in continuous flow synthesis, kinetic and thermodynamic analysis and unit operation development, but accurate technoeconomic analyses have paramount importance on economic viability. The literature abounds with societally critical Active Pharmaceutical Ingredients (APIs) whose flow synthesis has been demonstrated at lab/pilot scale (Mascia et al., 2013). Nevertheless, very few studies address the identification of promising APIs for CPM implementation and economic implications (Jolliffe and Gerogiorgis, 2015a-b, 2016).

This paper presents a systematic multi-parametric statistical study of a pharmaceutical trade dataset (PRODCOM) to identify key candidates for viable CPM implementation. Product ranking, trend plotting and hypothesis testing are used to explore and establish added value correlations for many drug substance (API) and drug formulation products.

Pharmaceutical product demand undergoes seasonal as well as long-term fluctuations due to a nexus of interweaved biological, environmental, business and political causes. As technoeconomic criteria are scarcely used, heuristic rules are routinely employed in order to determine low-risk priorities for pharmaceutical R&D, particularly in CPM: the principles of simplicity, efficiency, robustness and added value are hard to quantify. Limiting the number of continuous flow synthesis steps is another obvious policy, but that may often imply the need to procure high-MW feedstock molecules by outsourcing. Limiting the number of intermediate/final (esp. multiphase) separations is also pursued.

Accordingly, no systematic economic evaluation has been conducted to identify the most promising APIs for production-scale CPM implementation, beyond the possibility of empirical or semi-quantitative ranking of priorities (Jolliffe and Gerogiorgis, 2015a). This work presents an economic study and statistical analysis of pharmaceutical trade data in order to rank various products and thus identify the strongest CPM candidates. Physicochemical properties and API molecular complexity have also been compared toward understanding the connection between chemical structure and economic impact.

## 2. Pharmaceutical trade data: Methodology and analysis

This paper presents a multi-parametric statistical study of an enormous pharmaceutical trade dataset, aimed at identifying viable API candidates toward CPM implementation. The UK ONS PRODCOM (PRODucts of the European COMmunity) database (Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations) is the basis of our study, as compiled from UK manufacturers' sales and trade (2008-2013). Goods are classified as per EU Combined Nomenclature (Official EU Journal, 2014); products are divided into 48 groups (25 for drug substances and 23 for formulations). The dataset comprises statistics for the total volume, value (cost) and average price of imports (IMP), exports (EXP) and the trade balance (TRB) (i.e. exports minus imports).

### *2.1. Trade balance evolution trend plots*

The analysis of top market products has been carried out by considering volume (tons), value (GBP) and price (GBP/ton) data of the total exports, imports and trade balance. Products have been ranked according to these nine key trade variables by computing an overall score which is defined as the sum of ranks of each product in all nine categories. Product ranks emerge via score classification and can be plotted versus time (Fig. 1).

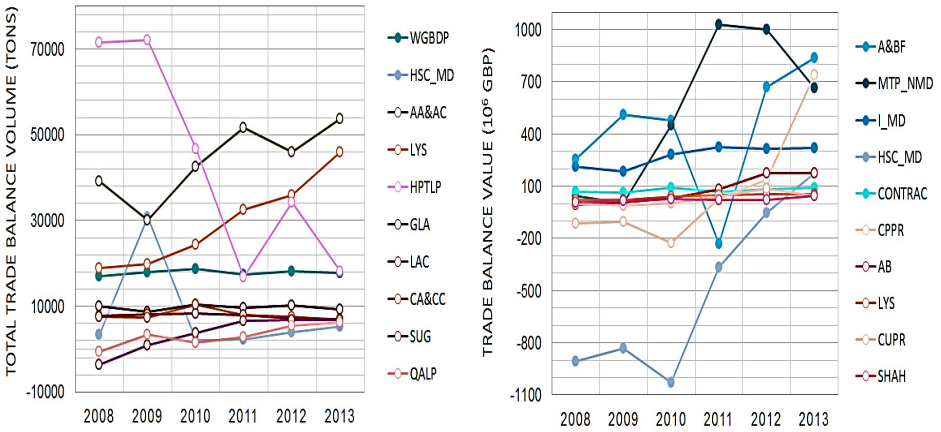


**Figure 1.** Top 10 products as ranked by total trade balance volume (left) and value (right).

*2.2. Statistical hypothesis testing*

Hypothesis testing has been conducted to probe correlations between substance and formulation products, via the Mann-Whitney-Wilcoxon (MWW) test (Sprent, 1998). The ranks of the 25 basic substances (APIs) and of the 23 pharmaceutical formulations have been compared for each year individually: the MWW test null hypothesis ($H_0$) is the similarity between rankings of basic substances and pharmaceutical formulations. The shape of the distributions has also been tested via the Kolmogorov-Smirnov test.

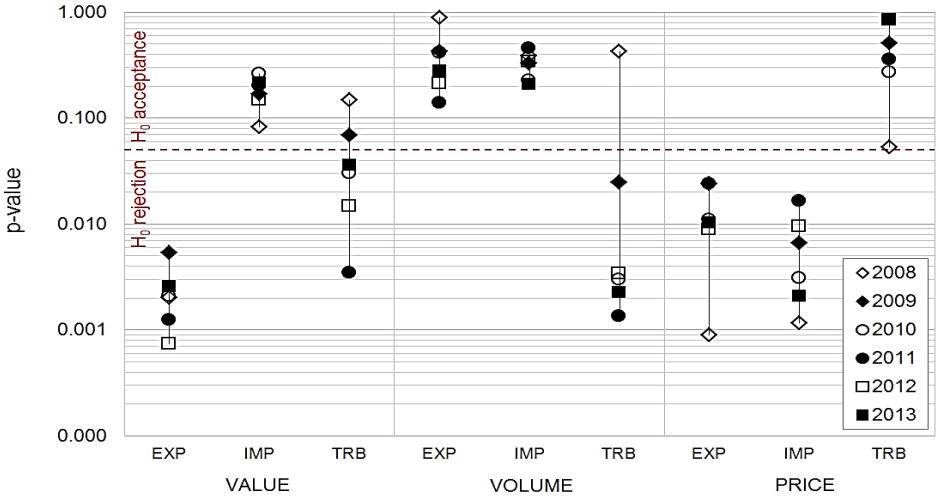


**Figure 2.** Mann-Whitney-Wilcoxon (MWW) test results for all nine parameters and all years.

*2.3. Bubble diagrams*

To probe the relation between physicochemical properties and economic characteristics, basic substance (API) pharmaceutical products have been further studied; functional groups of more than 100 pharmaceutical compounds (Official EU Journal, 2014) have been correlated with key molecule properties from a vast database (PUBCHEM, 2015). Molecular size and structure are quantified by molecular weight (MW) and the Bertz-Hendrickson-Ihlenfeldt complexity (CPX) formula (Bertz, 1981; Hendrickson, 1987). Bubble diagrams of the top 10 products have been plotted for MW, CPX and CPX/MW.

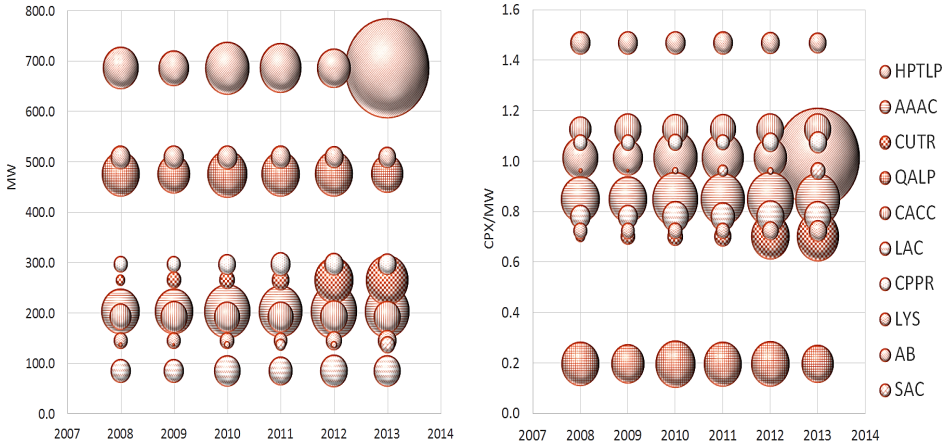


**Figure 3:** Top 10 products as ranked by total export volume vs. MW (left) and CPX/MW (right).

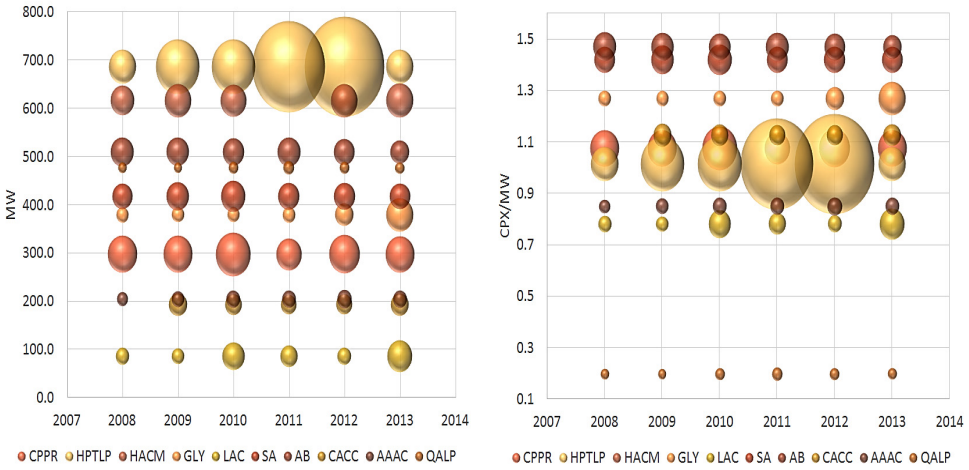


**Figure 4.** Top 10 products as ranked by total export value vs. MW (left) and CPX/MW (right).

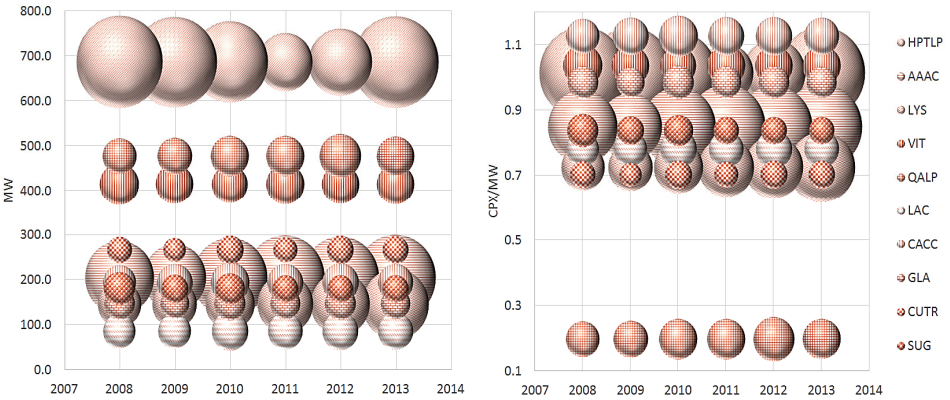


**Figure 5:** Top 10 products as ranked by total import volume vs. MW (left) and CPX/MW (right).

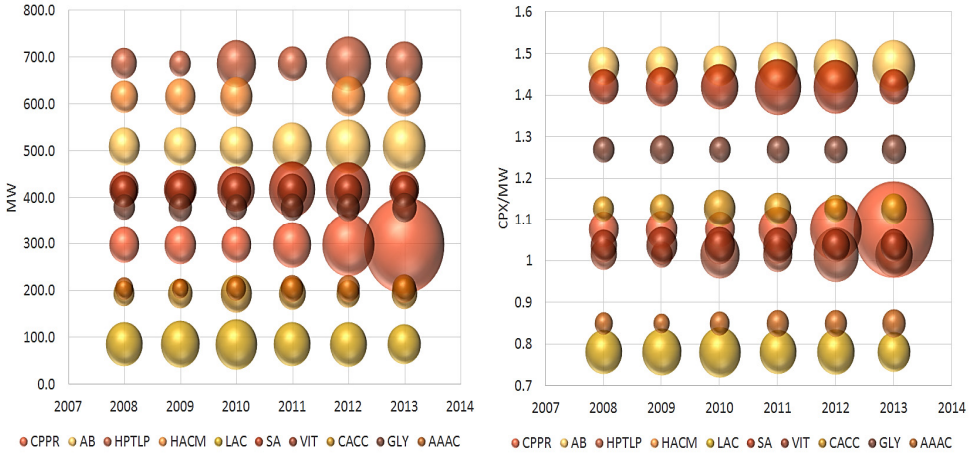


**Figure 6.** Top 10 products as ranked by total import value vs. MW (left) and CPX/MW (right).

## 3. Results and discussion

*3.1. Trade balance evolution trend plots*

The *trade balance volume* trend plot (Fig. 1L) is dominated by eight substance codes: acyclic amides and carbamates (AAAC), lysine derivatives (LYS), hormones, prostaglandins, thromboxanes, leukotrienes and polypeptides (HPTLP), glutamic acid and salts (GLA), lactones (LAC), cyclic amides and carbamates (CACC), sugars and derivatives (SUG) and quaternary ammonium salts (QALP). The strongest growth trend is observed for AAAC and LYS; the remaining six show slower growth or even decline. Coated bandages (WGBDP) and hormone/corticosteroid preparations (HSC_MD) are the only two formulations which appear in the top 10 product codes by trade balance.

The *trade balance value* trend plot (Fig. 1R) includes several other product codes: this time, there is an equal split between substances and formulations in the top 10 codes. Antisera and blood fractions (A&BF), therapeutic (MTP_NMD), insulin (I_MD) and hormone/steroid (HSC_MD) doses and contraceptives (CONTRAC) dominate the top part; pyrimidine/piperazine ring compounds (CPPR), antibiotics (AB), LYS, unfused pyrazol ring compounds (CUPR) and heparin derivatives (SHAH) are found lower. Several of these codes (A&BF, CPPR, HSC_MD) display strong trade growth trends.

*3.2. Statistical hypothesis testing*

The null hypothesis $H_0$ (similarity of trends between drug substances and formulations) is accepted when the p-value (probability of observed or more extreme results under $H_0$) is smaller than the significance level (5% in this study), and rejected when it is larger. MWW test results computed are mostly consistent for 2008-2013 and shown in Fig. 2. Definite $H_0$ rejection is observed for import and export prices as well as export value, indicating a strong incentive to explore the economic impact of CPM implementation.

*3.3. Bubble diagrams*

*Export volume rankings* for both categories (Fig. 3) show the top 10 product codes: hormones, prostaglandines, thromboxanes, leukotrienes and polypeptides (HPTLP), acyclic amides and carbamates (AAAC), unfused triazine ring compounds (CUTR), quaternary ammonium salts (QALP), cyclic amides and carbamates (CACC), lactones (LAC), pyrimidine and piperazine ring compounds (CPPR), lysine derivatives (LYS), antibiotics (AB) and salicylic acid (SAC) thus have the highest commercial importance. *Import volume rankings* (Fig. 5) feature seven of these codes (an overlap is clear here) and additionally showcase that provitamins and vitamins (VIT), glutamic acid and salts (GLA) as well as sugars and derivatives (SUG) are highly sought and thus imported. Due to the high added value of the latter, opportunities for UK production are evident.

*Export value rankings* for both categories (Fig. 4) present another remarkable feature: beyond the seven previous codes (CPPR, HPTLP, LAC, AB, CACC, AAAC, QALP) which are again encountered to show key strengths of the UK pharmaceutical industry, the presence of human/animal blood and micro-organism culture products (HACM) indicates lower volumes but quite high prices due to vast added biotechnological value. Glycosides and vegetable alkaloids (GLY) as well as sulfonamides (SA) resonate this pattern, obviously due to their extreme therapeutic importance and high global demand. *Import value rankings* (Fig. 6) comprise only product codes already present in previous plots (CPPR, AB, HPTLP, HACM, LAC, SA, VIT, CACC, GLY, AAAC), confirming global trade overlaps and reaffirming the very strong demand for these product families. For most product codes, a remarkable trend is that trade volume and value increase over time not as much as function of MW, but of scaled molecular complexity (CPX/MW).

## 4. Conclusions

The statistical analysis of economic data indicates that pyrimidine and piperazine ring compounds, cyclic and acyclic amides and carbamates, lactones and sulfonamides emerge as the most promising small molecule candidates for continuous manufacturing. Antibiotics, hormones, amino acids and several polypeptides also present an even stronger business case in the formulations category. Clear increasing sale volume and value trends have been strongly correlated with molecular complexity and highlight attractive R&D opportunities which are under investigation (Robertson et al., 2015). Drug substances (APIs) have also been further studied via an extensive molecular complexity survey: physicochemical properties of hundreds of API molecules have been compiled, and the respective trade statistics have been plotted and analysed as a function of molecular weight (MW) and the Bertz-Hendrickson-Ihlenfeldt structural complexity formula (Bertz, 1981; Hendrickson, 1987). To the best of our knowledge, this is the first study which explores correlations between pharmaceutical trade statistics and molecular complexity in the context of continuous process systems engineering. The need for systematic statistical evaluation of candidates in advance of R&D for CPM is clearly established: the methodology followed can support flow chemistry and process R&D (Gerogiorgis and Jolliffe, 2015) by rapid screening of viable candidates. Prior instances of (or current efforts for) continuous flow synthesis schemes are critical, but their relative importance must be assessed by economic as well as technical metrics. Higher geographic, temporal and molecular resolution of production, demand and trade data will aid the proposed multi-parametric quantitative evaluation more effectively, by pinpointing the strongest market opportunities for business and CPM process R&D: the limited public-domain availability of sensitive datasets illustrates their high importance.

## References

N.G. Anderson, 2012, Using continuous processes to increase production, *Org. Process Res. Dev.* 16(5), 852–869.

S.H. Bertz, 1988, Branching in graphs and molecules, *J. Am. Chem. Soc.* 19(1-3), 65–83.

D.I. Gerogiorgis, H.G. Jolliffe, 2015, Continuous pharmaceutical process engineering and economics, *Chim. Oggi-Chem. Today* 33(6), 29–32.

J.B. Hendrickson, P. Huang, A.G. Toczko, 1987, Molecular complexity: A simplified formula adapted to individual atoms, *J. Chem. Inf. Comput. Sci.* 27(2), 63–67.

H.G. Jolliffe, D.I. Gerogiorgis, 2015a, Process modelling and simulation for continuous pharmaceutical manufacturing of ibuprofen, *Chem. Eng. Res. Des.* 97: 175–191.

H.G. Jolliffe, D.I. Gerogiorgis, 2015b, Plantwide design and economic evaluation of two continuous pharmaceutical manufacturing (CPM) cases: ibuprofen and artemisinin, *Comput. Aid. Chem. Eng.* 37: 2213–2218.

H.G. Jolliffe, D.I. Gerogiorgis, 2016, Process modelling and simulation for continuous pharmaceutical manufacturing of artemisinin, *Chem. Eng. Res. Des.*, in press (2016).

S. Mascia et al., 2013, End-to-end continuous manufacturing of pharmaceuticals: integrated synthesis, purification and final dosage formation, *Angew. Chem. Int. Ed.* 52(47), 12359–12363.

Official E.U. Journal, 2014, L 312, 57, 1–923, European Union, Brussels (31 October 2014).

K. Plumb, 2005, Continuous processing in the pharmaceutical industry – changing the mind set, *Chem. Eng. Res. Des.* 83(A6), 730–738.

PRODCOM Statistical Database, 2015, Office for National Statistics (ONS), London, UK.

PUBCHEM Statistical Database, National Institutes of Health (NIH), Bethesda, MD, USA.

J. Robertson et al., 2015, *Continuous Pharmaceutical Manufacturing of Antibiotics*, Process Design Project Report, Disc. of Chem. Engineering, University of Edinburgh, UK.

S.D. Schaber et al., 2011, Economic analysis of integrated continuous and batch pharmaceutical manufacturing: a case study, *Ind. Eng. Chem. Res*., 50(17), 10083–10092.

P. Sprent, 1998, *Data Driven Statistical Methods*, Chapman and Hall/CRC Press, London, UK.